CLAIMS

1. A method for creating a unique DNA genome database, comprising the steps of:

providing available genomic sequence data in a first database

enumerating regions of identity between genomic sequences in the first data base and other genomic sequences in the first database as if the first database was also a query database, the enumerating being done on a computer having a processor, memory, input/output mechanisms; and

removing from the first database, genomic sequences that are nearly identical to a region of a longer genomic sequence, whereby a unique DNA genome database is created.

2. A system for creating a unique DNA genome database, comprising;

means for providing available genomic sequence data in a first database

means for enumerating regions of identity between genomic sequences in the first data base and other genomic sequences in the first database as if the first database was also a query database, the enumerating being done on a computer having a processor, memory, input/output mechanisms; and

means for removing from the first database, genomic sequences that are nearly identical to a region of a longer genomic sequence, whereby a unique DNA genome database is created.

3. A computer program for finding near identities in a DNA sequence database, comprising:

a first code mechanism for comparing a DNA sequence on a query database to DNA sequences on a data database, wherein a tag array (designated as Qtags) is generated for each of the DNA sequences on the query database and wherein a tag array (designated Dtags) is generated for each of the DNA sequences on the data database; and

a second code mechanism for comparing each Qtag to each Dtag, using a comparison model, wherein near identities of sequences in the two databases are identified.